**Does Weather impact Food sales?**

**Abstract:**

Weather is a major factor in food chain, Food production requires sunlight, rain, wind and natural manure to be sufficient for the crops to be grown. This is where weather plays a major role in the quantity of food product during the phase of a year. These changes in the climate can be a significant factor involved in contribution to the quantity produced and this in turn reflects on the price of the food. The cheaper the price the more it will be sold, the quantity and the quality of the products produced makes the cost of the product.

Weather data can be used to identify the food selection. Food is always considered to be an industry with high and a constant revenue, with that in the mix people like to choose their food based on the weather on a particular day. This is an interesting analysis which can be perform to prepare food based on the forecasted weather result to increase the sale on any given day by displaying the dishes most suited on the weather. Bayesian probabilistic contentions are utilized to determine romanticized conditions for finding the best investigation for numerical climate forecast. With respect to future environmental change, the Climate Risk Index can fill in as a notice flag demonstrating past weakness which may additionally increment in areas where extraordinary occasions will get to be more regular or more serious through environmental change.

**Contribution:**

The previous decade has seen the blossoming of a few novel methods for climate figure check and measurements for atmosphere display assessment. In spite of being centered around tending to the requirements of particular (climate or atmosphere) applications, a large portion of these strategies can be imagined and adjusted to serve in a multidisciplinary setting. The commitments on factual methodologies for both climate and atmosphere show assessment (and related applications), with the trust of improving logical trade and cooperative energies between the climate and atmosphere groups. There are major components making up a wild fire in the environment, the most important factor is weather, which is continuously changing. In this paper, we deal with the role of weather spreading of wildfires. Several elements of weather are considered and it directly impacts in managing the burns and fighting fires.

Weather models very commonly used on a day-to-day basis. Using the past and the current weather data observed, forecast the future movement of the weather. Weather forecast concentrates on taking current perceptions of climate and handle these information with the help of computer models to calculate the future condition of climate. There are various factors affecting the change in weather few of which is as follows Temperature, Humidity, Pressure, Dew point, Wind speed, Wind direction, Visibility. All the above mentioned variables serve as input for the model.

Climate research plays a key role in understanding the challenge posed by climate and in developing the necessary solutions to benefit the environment, society and the business sector. The raise in the temperature causes global warming which in turn leads to the raise in the sea water level. The greenhouse effect or raise in the carbon di oxide also influences directly in the raise of global temperature. To predict the weather it is essential to identify all the factor influencing the changes.

The guideline of concordance amongst Climate and the climate depends on reasonable experience. It might appear to negate standards expressed somewhere else however the reality remains: Climateand the four seasons have distinctive effect on the human body. Climate turns out to be a piece of the body in the wake of being devoured however the four seasons (that is the natural variable) dependably impacts remotely on the body. The Chinese dietary theory recommends that you grasp your local nourishments notwithstanding eating privately developed sustenances and those in season. What is unfortunate about western eating routine is that nourishments are again and again synthetically treated as opposed to being developed normally. Common, home-developed and concoction free items are generally nutritious.

Incomprehensibly, hot dishes spiked with chiles or ginger can really chill you off in hot climate. Here, hot Thai steak (left) and more hot-however cooling formulas from gourmet specialists around the globe.

Climbing in cool climate tends to smolder a few hundred more calories for each day than in the milder temperatures of the spring or fall. Our bodies work overtime to maintain the interior warmth expected to survive [source: Braaten]. In any case, for some odd reason, most individuals don't get in shape in the winter - they pick up it. It's no mystery why this happens.

Most by far of the weather doesn't go on icy climate climbs every day. They work in warmed structures where doughnuts lie temptingly in the lounge; they spend a larger piece of their days in warm homes where relaxation exercises like sitting in front of the TV run as an inseparable unit with nibbling. What's more, they assemble at occasion parties and other get-togethers where plunges, sweets and fatty beverages proliferate. That doesn't need to be you.

A hot bowl of oatmeal is completely what you make it. Oats are characteristic and stuffed with fiber. Mixed with sweet solidified organic products, it becomes an effective wellspring of vitamins and cancer prevention agents. Two-percent or skim milk offers calcium without pointless fat. Furthermore, it doesn't need to end there. Two tablespoons of ground flax seed can help bring down cholesterol. Sprinkle some almonds or walnuts on top for included sustenance and surface. Every last bit of it together makes a tremendous wellspring of sound fuel to move you as the day progressed.

Americans consume more than 36 million metric huge amounts of meat and poultry annually. Livestock and poultry represent over portion of U.S. farming money receipts, regularly over $100 billion for each year. Changes in climate could influence animals both straightforwardly and in a roundabout way. Dry spell may undermine field and sustain supplies. Dry season lessens the amount of value search accessible to brushing domesticated animals. Some regions could encounter longer, more serious dry spells, coming about because of higher summer temperatures and diminished precipitation. For animals that depend on grain, changes in yield generation because of dry season could likewise become a problem.

**Work by Competitor:**

Wildfire behavior and its severity lead from the combinations of many factors, whereas weather is one of the factors among it. It could be caused by major natural ignitions and weather is one of the compounds which is used to predict the wildfire. Regional weather patterns and climatic change can increase risk the behavior of the wildfires. The frequency of large wildfires and the total area burned have been consistently expanding in the Western United States, with global warming being a major contributing factor it is expected that the factors contributing to weather plays a major role in wildfire.

The frequency of large wildfires and the total area burned have been steadily increasing in the Western United States, with global warming being a major contributing factor. Firefighters know about the almost unusual nature of woods flames. Be that as it may, analysts now say they could be better arranged. Following quite a while of advancement, Janice Coen and her associates at the National Center for Atmospheric Research are get ready to dispatch innovation that incorporates information on climate, geology and different variables to anticipate how fires spread in a way already incomprehensible. The data represents only 40% for worldwide vitality use. Therefore the data set can be skewed to some extent making a biased contribution.

The competitor used the methods carefully and designed the techniques in such a way that the system gives the best result over the course and the model doesn’t over fit the requirement. But to out perform the methodology used by my competitor I have improvised the ETL (Extract Transfer Load) to the optimized situation that it pulls all the data periodically and displays the result real time situation, Spotfire has been used as a dashboard tool to visualize the graph real time to provide filters for the variables (independent variables) and summarize the results to the end users to simplify the process and act to the needs.

Late methodological and reasonable advancements in measurable downscaling. We particularly welcome commitments tending to spatial-transient and multi-variable changeability (specifically of extraordinary occasions); the advancement of factual models for sub-every day fluctuation, for example, convective occasions; the joining of process comprehension into inclination revision techniques; the choice of indicators to catch environmental change; the execution and included benefit of downscaling strategies regular to centennial scales (counting the capacity to extrapolate past watched values); the improvement of process-based approval diagnostics for measurable downscaling; and the evaluation of favorable circumstances and intrinsic constraints of various methodologies.

**About The Data**

The data is obtained for the city of Aarhus in Denmark and Brasov in Romania. The data is calculated on a random time in a day to identify the variation across the day and at the end the mean value is consider for during that time period. The data includes values from all the 7 days during the time period. Multiple records are noted throughout the day to accurately identify the range. The variables are calculated in different denominations such as [Dew point](http://en.wikipedia.org/wiki/Dew_point) in degrees Celsius, Humidity (percentage), Pressure in mBar, Temperature in degrees Celsius, Wind direction in degrees and Wind speed in kilometers per hour (kph).

Data Structure

The data consists of the following columns

1. Temperature
2. Humidity
3. Pressure
4. Dew point
5. Wind speed
6. Wind direction
7. Visibility
8. Timestamp

Let’s look at the description of all these variables in depth below.

Temperature :

Earth is getting warmer; the cause could be human activity or natural variability. Since the beginning of the industrial revolution, it is observed as an evidence of risen temperature. Of course, the temperature fluctuates by many degrees everyday where we live, by accumulation of this fluctuation now had accrued and lead and to be in global warming. The global temperature record indicates an average over the entire surface of the planet. A one degree global change is significant as because it takes a huge amount of heat to warm all oceans, atmosphere and land. Since the year 2000, land temperature changes are 50% greater in the US than ocean temperature changes; two to three times greater in the Arctic Ocean and Antarctic Peninsula, warming is more in the Arctic Ocean, and the next is in the Indian and Pacific oceans and the third largest in the most of Atlantic Ocean. The global temperature changes mainly depends on how much energy the planet receives from the sun and how much it radiates back.

Humidity :

It tells the moisture content of the atmosphere. When the humidity is high it feels oppressive outside because of sweat doesn’t evaporate. When the humidity is low you feel cooler but moisture evaporated. The relative humidity is dependent on air temperature. If the water vapour content stays the same and the temperature drops the relative humidity increases.

There are three fundamental estimations of moistness: total, relative and particular. Outright dampness is the water substance of air at a given temperature communicated in gram per cubic meter.[2] Relative moistness, communicated as a percent, measures the flow total stickiness with respect to the most extreme (most elevated point) for that temperature. Particular mugginess is a proportion of the water vapor substance of the blend to the aggregate water content on a little premise.

Pressure :

Air gets lighter with increasing altitude. Atmospheric pressure is around 1014 mill bars at sea level. At sea level, the air that surrounds presses down the bodies at 14.5 pounds per square inch.

Pressure (symbol: p or P) is the [force](https://en.wikipedia.org/wiki/Force) applied perpendicular to the surface of an object per unit [area](https://en.wikipedia.org/wiki/Area) over which that force is distributed. [Gauge pressure](https://en.wikipedia.org/wiki/Gauge_pressure) (also spelled gage pressure)[[a]](https://en.wikipedia.org/wiki/Pressure" \l "cite_note-1) is the pressure relative to the ambient pressure.

Dew Point:

Dew point may drop by some possible influences; as when the temperature drops it’s enough to get fog and dew, water condenses out of the air and lowering the dew point. When it is hotter outside, the water will evaporate from the ground faster and the dew point raises. And when it is hottest outside, there is more wind and the mixing of low altitude and high altitude air lowers the dew point. Dew point is dependent on temperature. Dew point is a more reliable indicator of humidity because it is not changed by a change in air temperature and doesn’t fluctuate much throughout the day.

Wind Speed :

Wind, which is as called as air movement, is integral to all types of weather conditions. Wind speed is caused by air moving from high pressure to low pressure, usually due to changes in temperature. Wind speed is influenced by various elements and circumstances, working on shifting scales (from smaller scale to full scale scales). There are likewise connections to be found between wind speed and wind course, strikingly with the weight inclination and landscape conditions.

Wind Direction :

Wind direction is obtained by the [direction](https://en.wikipedia.org/wiki/Direction_(geometry)) from which it initiates. For example a wind coming from the south is shown as 180 degrees and the one from the east is 90 degrees. Air pressure, which is largely caused by differential heating of the air by the sun and ground conditions, controls the way air flows. Instruments can be utilized to quantify wind direction, for example, the windsock or a wind vane. Both of these instruments work by moving on air resistance. The way a climate vane is pointed by winning winds demonstrates the direction of wind direction.

Visibility :

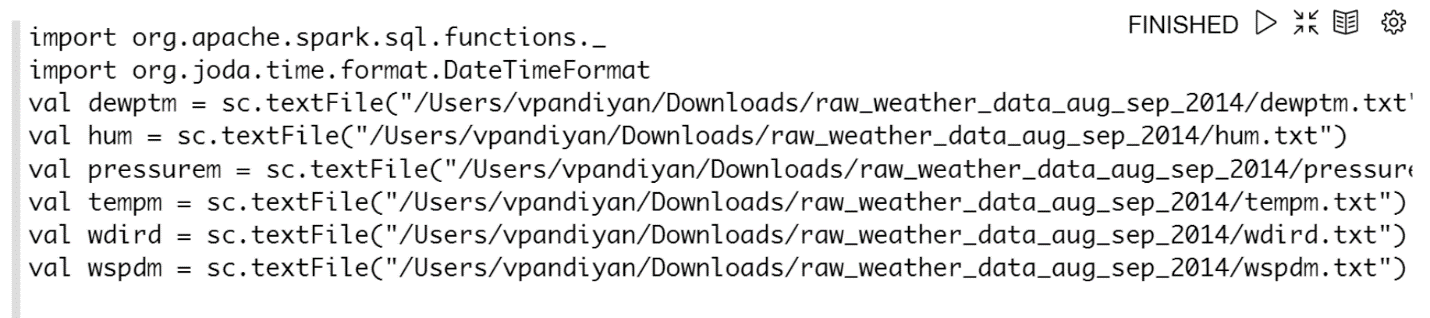
Visibility is a measure of the distance at which an object or light can be clearly seen . visibility affects all forms of traffic, aviation and sailing. Visibility to transparency of air; in dark visibility still the same as in daylight for the same air. If the temperature is warm, the atmosphere has a greater capacity to hold water in its vaporous state than if it is cold. Vapour can change into visible water in one other way.

Timestamp:

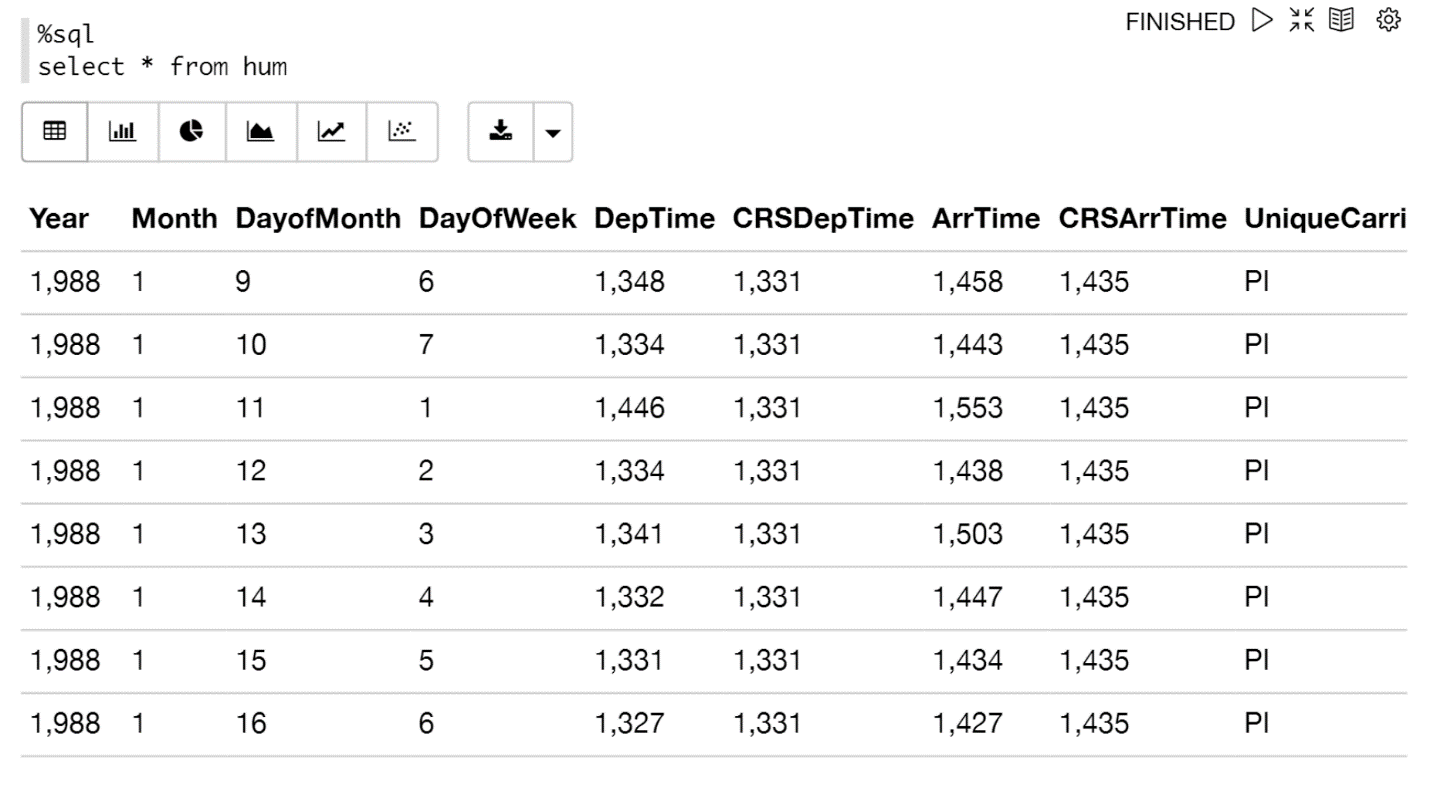
Timestamp records the time at which the values are recorded. It is of the form mm/dd/yyyy and it also provides the time at which the values are noted.

**Data imported with Zeppelin :**

Commands used for importing data into Zeppelin.

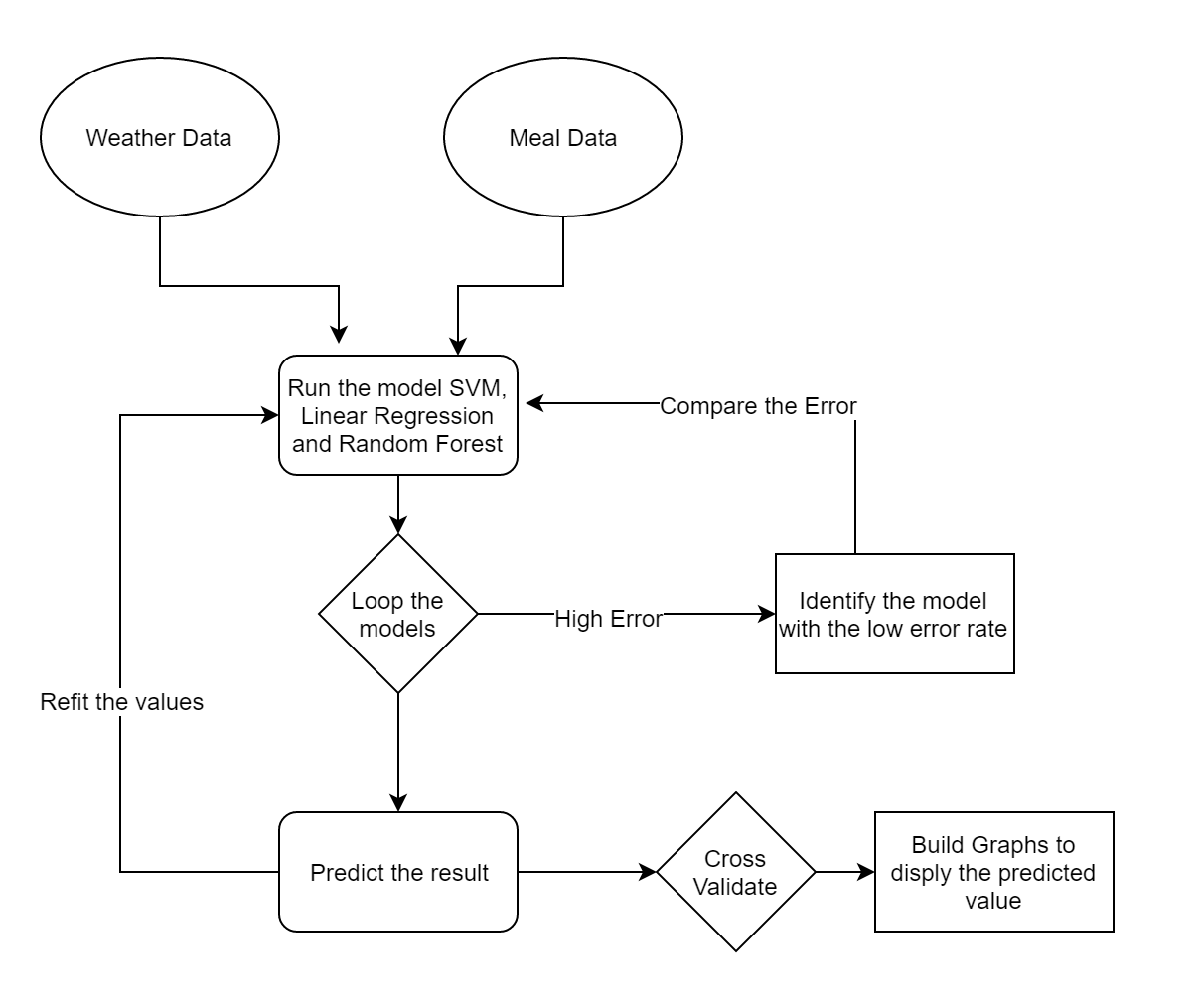


When displayed the table looks like.



**Methods:**

**All models flow the below flow chart.**

****

**Linear Regression:**

Simple straight relapse is a factual method that permits us to summarize and ponder connections between two consistent (quantitative) factors: One variable, denoted x, is viewed as the indicator, illustrative, or free factor. The other variable, denoted y, is viewed as the reaction, outcome, or ward variable.

Since alternate terms are utilized less every now and again today, we'll utilize the "indicator" and "reaction" terms to allude to the factors experienced in this course. Alternate terms are mentioned just to make you mindful of them should you experience them in different fields. Simple direct relapse gets its descriptor "simple," since it concerns the investigation of just a single indicator variable. Interestingly, multiple direct relapse, which we concentrate later in this course, gets its descriptor "multiple," on the grounds that it concerns the investigation of two or more indicator factors.Before continuing, we must elucidate what sorts of connections we won't think about in this course, namely, deterministic (or practical) connections. Here are some examples of other deterministic connections that understudies from past semesters have shared. For each of these deterministic connections, the condition precisely depicts the relationship between the two factors. This course does not examine deterministic connections. Rather, we are occupied with measurable connections, in which the relationship between the factors is not great.

Here is an example of a measurable relationship. The reaction variable y is the mortality because of skin growth (number of passings per 10 million individuals) and the indicator variable x is the scope (degrees North) at the focal point of each of 49 states in the U.S. (skincancer.txt) (The information were compiled in the 1950s, so Alaska and Hawaii were not yet states. What's more, Washington, D.C. is incorporated into the informational index despite the fact that it is not actually a state.) skin malignancy versus state scope plot. You might envision that in the event that you lived in the higher scopes of the northern U.S., the less presented you'd be to the harmful beams of the sun, and in this manner, the less hazard you'd have of death because of skin tumor. The scramble plot backings such a theory. There gives off an impression of being a negative direct relationship amongst scope and mortality because of skin tumor, however the relationship is not great. To be sure, the plot displays some "pattern," yet it likewise shows some "diffuse." Therefore, it is a factual relationship, not a deterministic one.

Tallness and weight — as stature increments, you'd anticipate that weight will increment, however not flawlessly. Liquor consumed and blood liquor content — as liquor consumption increments, you'd expect one's blood liquor substance to increment, yet not impeccably. Essential lung limit and pack-years of smoking — as amount of smoking increments (as measured by the number of pack-years of smoking), you'd expect lung work (as evaluated by imperative lung limit) to diminish, however not flawlessly. Driving velocity and gas mileage — as driving pace increments, you'd anticipate that gas mileage will diminish, yet not consummately.

**Service Vector Machine:**

Support vector machine (SVM) is a supervised machine learning algorithm and is similar to the neural network family were there is an imaginary line drawn cross all the observations to minimize the error and maximize the prediction result.

**Pros:**

* It works truly well with clear margin of division
* It is powerful in high dimensional spaces.
* It is powerful in situations where number of dimensions is more prominent than the number of samples.
* It utilizes a subset of preparing focuses in the choice capacity (called bolster vectors), so it is likewise memory productive.

**Cons:**

* It doesn't perform well, when we have substantial informational collection in light of the fact that the required preparing time is higher
* It likewise doesn't perform exceptionally well, when the informational index has more commotion i.e. target classes are covering.
* SVM doesn't specifically give likelihood estimates, these are figured utilizing a costly five-overlay cross-approval. It is connected SVC method of Python scikit-learn library.

**Random Forest:**

Random Forest model does not overfit. You can keep running the same number of trees as you need. It is quick. Radom Forest method is developed based on a lot of decision tree method combined, It is also developed by random selection of variables. The result of the random forest method is based on many trees produced by the dependent variables. For example, let us consider 3 subsets, each subset can be of different size and different length. Running on an informational index with 50,000 cases and 100 factors, it delivered 100 trees in 11 minutes on a 800Mhz machine. For substantial informational collections the major memory requirement is the capacity of the information itself, and three whole number exhibits with the same dimensions as the information. In the event that proximities are figured, capacity requirements develop as the number of cases times the number of trees.

* It is unexcelled in precision among current algorithms.
* It runs productively on huge information bases.
* It can deal with a large number of information factors without variable cancellation.
* It gives estimates of what factors are important in the order.
* It creates an interior fair-minded estimate of the speculation blunder as the woodland building advances.
* It computes proximities between sets of cases that can be utilized as a part of bunching, finding exceptions, or (by scaling) give intriguing perspectives of the information.

All the techniques are properly measured to assess climate and atmosphere extremes in atmosphere display recreations. The mix of atmosphere model with observational and reanalysis items requires creative examinations and measurements of execution that make utilization of data identified with procedures producing inside atmosphere changeability and opening up climate and atmosphere extremes.

**Using these models:**

* The methodology will produce a fine resolution weather forecasting and the forecasting will provide useful information on the natural calamities especially wildfire.
* The model is designed in such a way to produce the results n timely manner.
* It increases the end result and creates awareness for the forest officials on a possible wildfire.
* The model can be linked to a device to automatically provide an alert in real time when the model predicts a possibility of a wildfire.

**Results:**

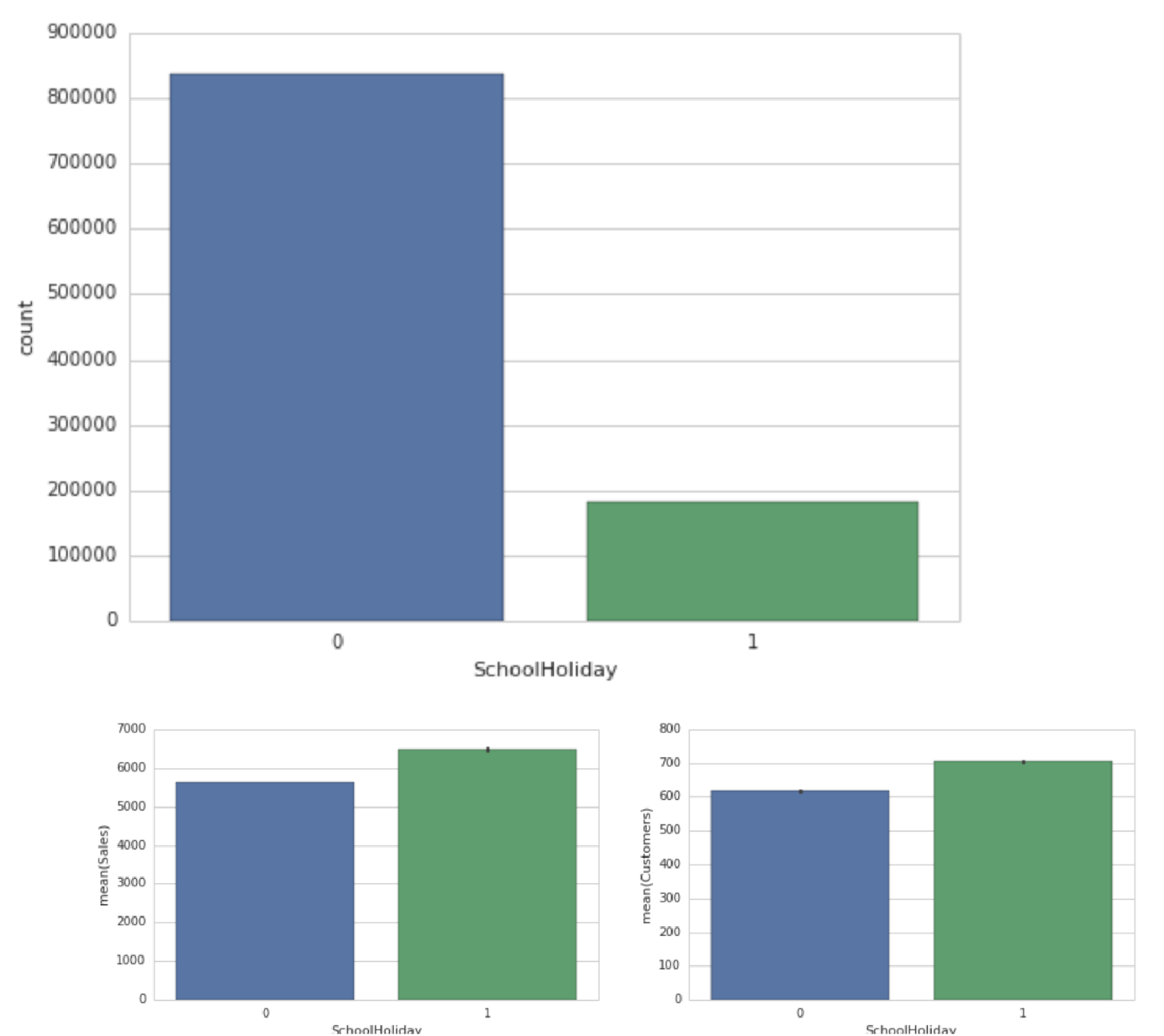
Tracking the variables provided during the training set from both sales and the weather we will be able to understand the correlation between them and considering the fact that the data talks about the sale on a given period of time, Weather can be a major influencing factor, all the techniques are properly measured to assess climate and atmosphere extremes in atmosphere display recreations. The mix of atmosphere model with observational and reanalysis items requires creative examinations and measurements of execution that make utilization of data identified with procedures producing inside atmosphere changeability and opening up climate and atmosphere extremes.

The results from the below methods shows the correlation between every variables and its sizeable impact on the food sales. The key variables identified by performing various techniques and the graphs below identifies the correlation between them.

Linear Regression trees are used when variables are continuous but the classification trees are only used when variables are categorical. In linear regression we try to establish if there is a relationship between dependent variables and independent variables. If the value of the variable increase, then we will check if the other variable increase or decrease. For example, let us consider Income and spending, In this case will can analyze if the income increases does the spending also increase or decrease. Using the metrics from the above method we can try to fit the model to the result dataset and predict future values. This helps use to forecast the future values for the same set of independent variables.

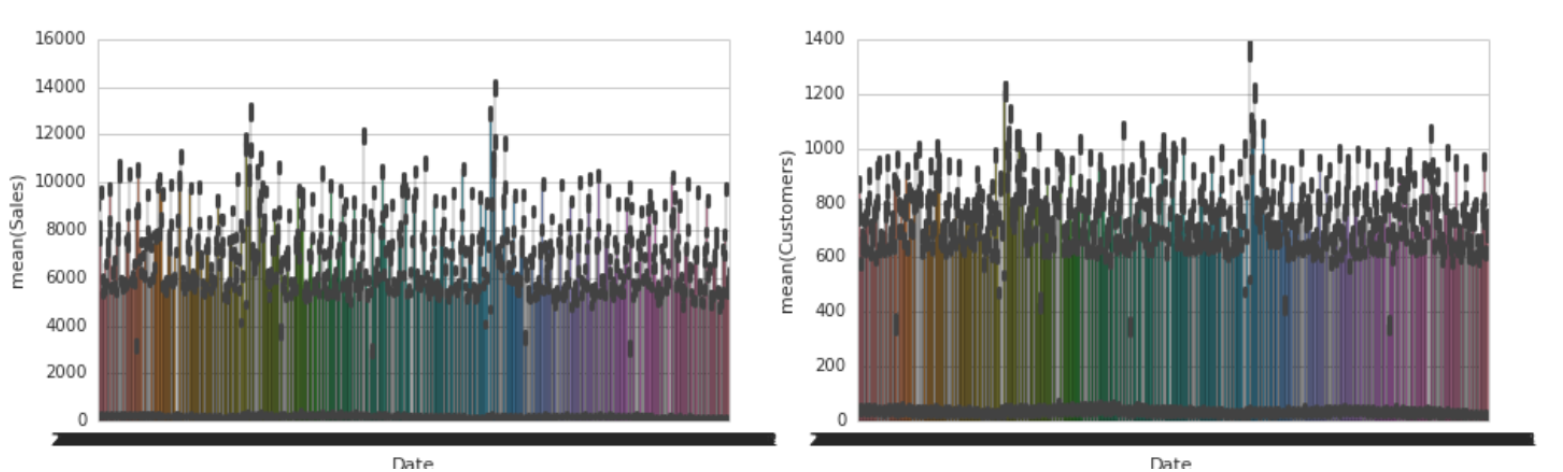
In our model we have used the methods to analyze the dependent variable in our case the sales of the product in different stores to the weather and the customer as independent variables. In the graph below we can see how the customer changes according to whether it is a school holiday or not.

Though overall count of the data is higher on a non-school holiday, the average sales and the avg customer during the holiday is higher compared to the non-holiday. This gives us an idea to further enhance the work to dig deep in to the holiday part of the data set.



The graph indicated the co relation between school holidays and the food sales.

To divide the analysis in to the next level we have divided the value of the customers and sales over a period of time. The Mean sales to the customer changes day to day ad filtering out the odd days we get the information on the factor that influenced the sales for that period of timeframe. Highlighting on the odd days shows us the information required to further examine the result for that period and we will also be able to track the reason for the odd increase or decrease in sales and the customer.



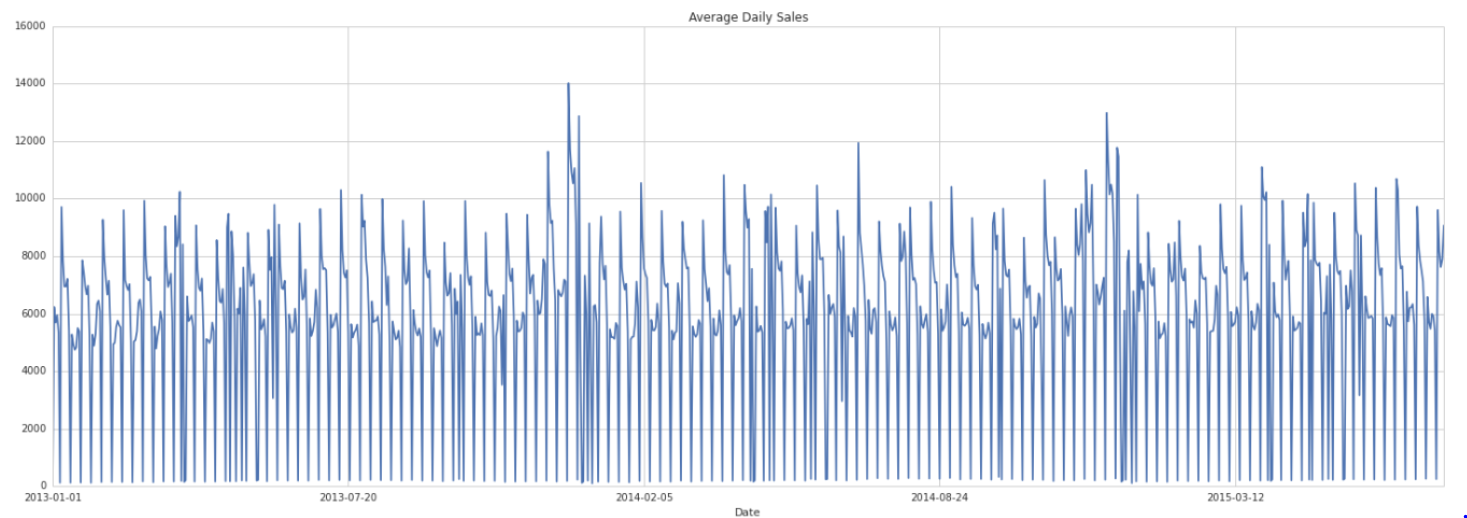
Radom Forest method is developed based on a lot of decision tree method combined, It is also developed by random selection of variables. The result of the random forest method is based on many trees produced by the dependent variables. For example, let us consider 3 subsets, each subset can be of different size and different length. Furthermore, these subsets can be overlapped by one another and If there are X1 to Xn independent variables it may have been in the first go that it took X1 X2 .. as a variable set and in the second go it took x3,x4,… as another variable set essentially we have selected the random subset of data and we have also selected a complete random set of variables.

To give an example.

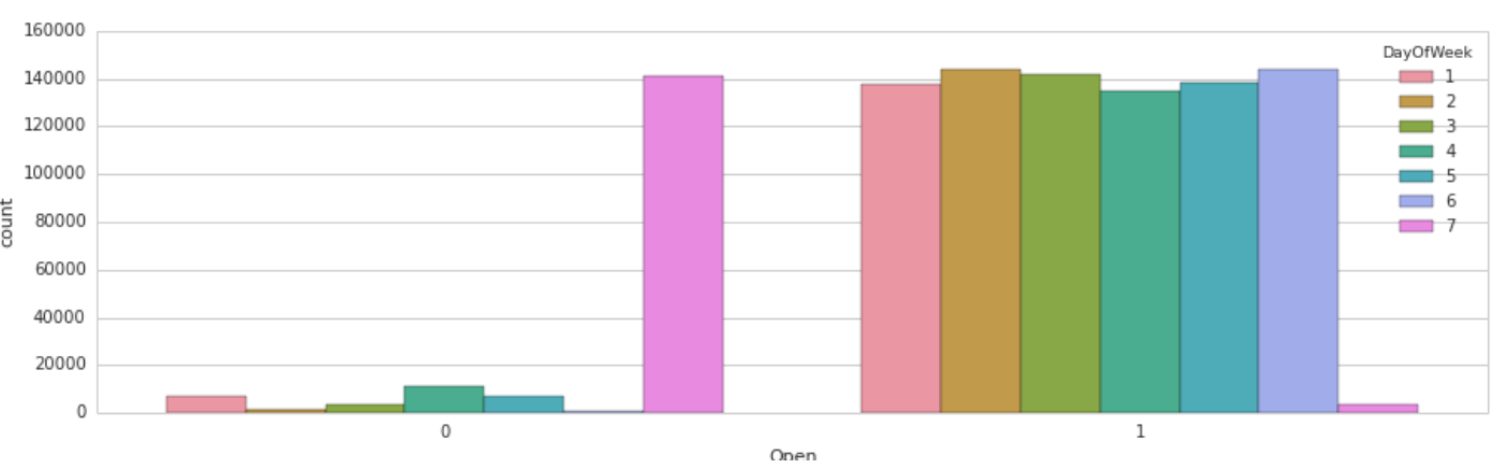
Let us consider 30 students in a class irrespective of their gender and there are 3 different variables / parameters Gender, Class and height. Out of all the students only 15 play football and now I would want to predict which student would play football as a hobby? In this problem we divide the students based on the sport that they play as hobby from the input. This is where we use decision tree model, It will divide the students based on all the variables in our case there are 3 variables defined.

This creates a homogeneous group of students based on the inputs provided, Gender can be a good variable to identify the best set of or group of students who is likely to play football.

In my case study, I have used the weather data to identify the set of people who are likely to purchase a food item on a given day. The variation provided in the model will identify the odd day on which there are a certain set of people who are likely to buy the product this variable are properly measured to indicate that it can be a promotion sales on that day.



This gives us a possibility to analyze the mood swings of the customer and we will also be able to track the seasonal distribution of the sales and the impact of seasonality on various occasions. The graphs give us a real picture of the various factor influencing the change in the customer mood and if we can track the sentiment analysis from twitter data we will be able to understand the emotional situation of the customer and the factors such as online reviews, comments and type of emotions running through the customers on a given period of time. The weather can always be an influencing factor when it comes to mood swings, From the data we can also find the variance with respect to day and night time for the customers to sales ratio to understand when are the customers likely to purchase large amount of products.



The above graph suggests the count of the customers coming on a weekday compared to a weekend and when the store is closed, from this example we are able to understand that there are a lot of customer not able to purchase on Sunday mainly because of the early close and it says that the customer missed is likely to be more than the customers on the weekdays.

**Discussion:**

When a product comes to market the demand and supply of the product determines the price of the product, this can be determined in two types; Processed and un-processed food. Processed food are something which isn’t grown directly or it is a derived product of the un-processed food. Un-processed food are grown from the field and should be directly taken in the raw form.

Environmental change is probably going to influence nourishment security at the worldwide, provincial, and nearby level. Climate change can disrupt food availability, reduce access to food, and affect food quality. For instance, anticipated increments in temperatures, changes in precipitation designs, changes in outrageous climate occasions, and decreases in water accessibility may all result in reduced agricultural productivity. Increments in the recurrence and seriousness outrageous climate occasions can likewise intrude on sustenance conveyance, and coming about spikes in nourishment costs after extraordinary occasions are relied upon to be more regular later on. Increasing temperatures can contribute to spoilage and contamination.

All the techniques are properly measured to assess climate and atmosphere extremes in atmosphere display recreations. The mix of atmosphere model with observational and reanalysis items requires creative examinations and measurements of execution that make utilization of data identified with procedures producing inside atmosphere changeability and opening up climate and atmosphere extremes.

**Conclusion**:

The methods above might reflect the same result as the competitor, but the method implied through this process is made in such way that the application interacts with the system periodicaly and live dashboards are generated as a result from the analysis, This gives the users the possibility to work with the data real time and understand and forecast the result real time to predict the sales on a particular day based on the weather data and it will give the competitor an edge to their competitors to work on the promotions. The results from the above methods shows the correlation between every variables and its sizeable impact on the food sales. The key variables identified by performing various techniques and the graphs below identifies the correlation between them.

The competitor used the methods carefully and designed the techniques in such a way that the system gives the best result over the course and the model doesn’t over fit the requirement. But to outperform the methodology used by my competitor I have improvised the ETL (Extract Transfer Load) to the optimized situation that it pulls all the data periodically and displays the result real time situation, Spotfire has been used as a dashboard tool to visualize the graph real time to provide filters for the variables (independent variables) and summarize the results to the end users to simplify the process and act to the needs.

For future research we can try to build a mobile application to give the customers an really time feeling of giving notifications about the situation of the market demand and price the products accordingly to satisfy the customers through promotions and mainly concentrate on the holiday season to give higher discounts to capture the market and make online market survey to understand the customer needs and perform a sentiment analysis to the reviews to improve the business further. The dashboard produced are real time and this will give our customer the experience of a real time understanding of the customers. To further the understanding the research can even look into the sentiment analysis on the reviews.